o Fitting a line to data

$$\text{Signal Model} \begin{cases} \text{Data}_i = y_i \\ = ax_i + b \\ \text{Noise} = r_i \text{ (or "}n_i\text{" — doesn't matter)} \end{cases}$$

$$\boxed{\text{Data} = \text{Signal} + \text{Noise}}$$

⟹ what makes this a probabilistic problem?

↳ noise!

$$p(r_i) = N(0, \sigma)$$

∴ $$p(y_i - s_i) = N(0, \sigma)$$

$$p(y_i) = N(s_i, \sigma)$$

⟹ $$p(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_i - s_i)^2}{\sigma^2}}$$

BUT s_i is now a LINE RELATIONSHIP not just the single "μ" as in the last lecture.

$$\ln L = \ln \left[\prod_i p(y_i) \right]$$

$$= \text{constant} - \frac{1}{2} \sum_i \frac{(y_i - a x_i - b)^2}{\sigma^2}$$

$$\underbrace{\qquad\qquad\qquad}_{\chi^2}$$

[NOTE] Also possible to modify the likelihood to reduce the impact of outliers (see the notebook).

○ <u>Goodness of Fit</u>

- MLE gives best-fit parameters of model.
- How do we know if the model is any good?

⇒ Max-likelihood should not be an unlikely occurrence ... need to know L distribution.

$$z_i = \frac{(y_i - s_i)}{\sigma_i}$$

$$\ln L = \text{constant} - \frac{1}{2} \sum_i z_i^2$$

$$= \text{constant} - \frac{1}{2} \chi^2 \quad \text{— chi-squared}$$

#data-points
#params

<u>Mean</u> of χ^2 distribution = $N - k$
<u>Std</u> of χ^2 distribution = $\sqrt{2(N-k)}$

χ^2 per degree of freedom

$$\boxed{\chi^2_{dof}} = \frac{1}{N-k} \sum_i z_i^2 \left.\right\} \begin{array}{c} \text{GOOD FIT} \\ \shortparallel \\ \chi^2_{dof} \approx 4 \end{array}$$

○ <u>Model Comparison</u> — much more on this
later in the course.

$$AIC_M = \underbrace{-2 \ln\left[L^{Max}(M)\right]}_{\propto \chi^2} + \underbrace{2k + \frac{2k(k+1)}{N-k-1}}_{\substack{\text{penalty on model} \\ \text{complexity.}}}$$

$\Rrightarrow$ Model with lowest AIC wins!

○ <u>Confidence Estimation</u> — Fisher matrix approach
is an approximation

$\Rrightarrow$ simulate $\boxed{NEW}$ data by resampling from
the data you have ... NOT MAGIC!!

<u>BOOTSTRAP</u> $\Rightarrow$ resample data (with replacement)
and compute statistics on each new
data realization ... <u>EASY</u>

<u>JACKKNIFE</u> $\Rightarrow$ remove one data point at a time,
and re-compute statistics
$\hookrightarrow$ gives $(N-1)$ new data
realizations ... <u>EASY</u>